METHODS AND COMPOSITIONS FOR THE IDENTIFICATION OF ANTIBIOTICS THAT ARE NOT SUSCEPTIBLE TO ANTIBIOTIC RESISTANCE

Related Applications

The present application claims priority from U.S. provisional patent application serial no. 60/393,237, filed on July 1, 2002, and U.S. provisional patent application serial no. 60/452,012, filed on March 5, 2003, which is expressly incorporated by reference.

Background of the Invention

Ribosomes are composed of one large and one small subunit containing three or four RNA molecules and over fifty proteins. The part of the ribosome that is directly involved in protein synthesis is the ribosomal RNA (rRNA). The ribosomal proteins are responsible for folding the rRNAs into their correct three-dimensional structures. Ribosomes and the protein synthesis process are very similar in all organisms. One difference between bacteria and other organisms, however, is the way that ribosomes recognize mRNA molecules that are ready to be translated. In bacteria, this process involves a base-pairing interaction between several nucleotides near the beginning of the mRNA and an equal number of nucleotides at the end of the ribosomal RNA molecule in the small subunit. The mRNA sequence is known as the Shine-Dalgarno (SD) sequence and its counterpart on the rRNA is called the Anti-Shine-Dalgarno (ASD) sequence.

There is now extensive biochemical, genetic and phylogenetic evidence indicating that rRNA is directly involved in virtually every aspect of ribosome function (Garrett, R. A., *et al.* (2000) *The Ribosome: Structure, Function, Antibiotics, and Cellular Interactions.* ASM Press, Washington, DC). Genetic and functional analyses of rRNA mutations in *E. coli* and most other organisms have been complicated by the presence of multiple rRNA genes and by the occurrence of dominant lethal rRNA mutations. Because there are seven rRNA operons in *E. coli*, the phenotypic expression of rRNA mutations may be affected by the relative amounts of mutant and wild-type ribosomes in the cell. Thus, detection of mutant phenotypes can be hindered by the presence of wild-type ribosomes. A variety of approaches have been designed to circumvent these problems.

One common approach uses cloned copies of a wild-type rRNA operon (Brosius, J., *et al.* (1981) *Plasmid* 6: 112–118; Sigmund, C. D. *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79: 5602–5606). Several groups have used this system to detect phenotypic differences caused by a high level of expression of mutant ribosomes. Recently, a strain of *E. coli* was constructed in which the only supply of ribosomal RNA was plasmid

encoded (Asai, T., (1999) *J. Bacteriol.* 181: 3803–3809). This system has been used to study transcriptional regulation of rRNA synthesis, as well as ribosomal RNA function (Voulgaris, J., *et al.* (1999) *J. Bacteriol.* 181: 4170–4175; Koosha, H., *et al.* (2000) *RNA.* 6: 1166–1173; Sergiev, P. V., *et al.* (2000) *J. Mol. Biol.* 299: 379–389; O'Connor, M. *et al.* (2001) *Nucl. Acids Res.* 29: 1420–1425; O'Connor, M., *et al.* (2001) *Nucl. Acids Res.* 29: 710–715; Vila-Sanjurjo, A. *et al.* (2001) *J. Mol. Biol.* 308: 457–463); Morosyuk S. V., *et al.* (2000) *J. Mol. Biol.* 300 (1):113-126; Morosyuk S. V., *et al.* (2001) *J. Mol. Biol.* 307 (1):197-210; and Morosyuk S. V., *et al.* (2001) *J. Mol. Biol.* 307 (1):211-228. Hui *et al.* showed that mRNA could be directed to a specific subset of plasmid-encoded ribosomes by altering the message binding site (MBS) of the ribosome while at the same time altering the ribosome binding site (RBS) of an mRNA (Hui, A., *et al.* (1987) *Methods Enzymol.* 153: 432–452).

Although each of the above methods has contributed significantly to the understanding of rRNA function, progress in this field has been hampered both by the complexity of translation and by difficulty in applying standard genetic selection techniques to these systems.

Resistance to antibiotics, a matter of growing concern, is caused partly by antibiotic overuse. According to a study published by the Journal of the American Medical Association in 2001, between 1989 to 1999 American adults made some 6.7 million visits a year to the doctor for sore throat. In 73% of those visits, the study found, the patient was treated with antibiotics, though only 5%-17% of sore throats are caused by bacterial infections, the only kind that respond to antibiotics. Macrolide antibiotics in particular are becoming extremely popular for treatment of upper respiratory infections, in part because of their typically short, convenient course of treatment. Research has linked such vast use to a rise in resistant bacteria and the recent development of multiple drug resistance has underscored the need for antibiotics which are highly specific and refractory to the development of drug resistance.

Microorganisms can be resistant to antibiotics by four mechanisms. First, resistance can occur by reducing the amount of antibiotic that accumulates in the cell. Cells can accomplish this by either reducing the uptake of the antibiotic into the cell or by pumping the antibiotic out of the cell. Uptake mediated resistance often occurs, because a particular organism does not have the antibiotic transport protein on the cell surface or occasionally when the constituents of the membrane are mutated in a way that interferes with transport of the antibiotic into a cell. Uptake mediated resistance is only possible in instances where the drug gains entry through a nonessential transport molecule. Efflux mechanisms of antibiotic resistance occur via transporter proteins. These can be highly specific transporters that transport a particular antibiotic, such as tetracycline, out of the cell or they can be more general transporters that transport groups

of molecules with similar characteristics out of the cell. The most notorious example of a nonspecific transporter is the multidrug resistance transporter (MDR).

Inactivating the antibiotic is another mechanism by which microorganisms can become resistant to antibiotics. Antibiotic inactivation is accomplished when an enzyme in the cell chemically alters the antibiotic so that it no longer binds to its intended target. These enzymes are usually very specific and have evolved over millions of years, along with the antibiotics that they inactivate. Examples of antibiotics that are enzymatically inactivated are penicillin, chloramphenicol, and kanamycin.

Resistance can also occur by modifying or overproducing the target site. The target molecule of the antibiotic is either mutated or chemically modified so that it no long binds the antibiotic. This is possible only if modification of the target does not interfere with normal cellular functions. Target site overproduction is less common but can also produce cells that are resistant to antibiotics.

Lastly, target bypass is a mechanism by which microorganisms can become resistant to antibiotics. In bypass mechanisms, two metabolic pathways or targets exist in the cell and one is not sensitive to the antibiotic. Treatment with the antibiotic selects cells with more reliance on the second, antibiotic-resistant pathway.

Among these mechanisms, the greatest concern for new antibiotic development is target site modification. Enzymatic inactivation and specific transport mechanisms require the existence of a substrate specific enzyme to inactivate or transport the antibiotic out of the cell. Enzymes have evolved over millions of years in response to naturally occurring antibiotics. Since microorganisms cannot spontaneously generate new enzymes, these mechanisms are unlikely to pose a significant threat to the development of new synthetic antibiotics. Target bypass only occurs in cells where redundant metabolic pathways exist. As understanding of the MDR transporters increases, it is increasingly possible to develop drugs that are not transported out of the cell by them. Thus, target site modification poses the greatest risk for the development of antibiotic resistance for new classes of antibiotic and this is particularly true for those antibiotics that target ribosomes. The only new class of antibiotics in thirty-five years, the oxazolidinones, is a recent example of an antibiotic that has been compromised because of target site modification. Resistant strains containing a single mutation in rRNA developed within seven months of its use in the clinical settings.

Summary of the Invention

The present invention provides compositions and methods which may be used to identify antibiotics that are not susceptible to the development of antibiotic resistance. In particular, rRNA genes from *E. coli* and other disease causing organisms are genetically engineered to allow identification of functional mutant ribosomes that may

be used as drug targets, *e.g.*, to screen chemical and peptide libraries to identify compounds that bind to all functional mutant ribosomes but do not bind to human ribosomes. Antibiotics that recognize all biologically active forms of the target molecule and are therefore not susceptible to the development of drug resistance by target site modification are thus identified.

The invention provides plasmid constructs comprising an rRNA gene having a mutant ASD sequence set forth in Figures 12, 13, 15, and 16, at least one mutation in the rRNA gene, and a genetically engineered gene which encodes a selectable marker having a mutant SD sequence set forth in Figures 12, 13, 15, and 16. The mutant SD-ASD sequences are mutually compatible pairs and therefore permit translation of only the mRNA containing the compatible mutant SD sequence, *i.e.*, translation of the selectable marker. In one embodiment, the selectable marker is chosen from the group consisting of chloramphenicol acetyltransferase (CAT), green fluorescent protein (GFP), or both CAT and GFP. In another embodiment, the DNA sequence encoding the rRNA gene is under the control of an inducible promoter.

The rRNA gene may be selected from a variety of species, thereby providing for the identification of functional mutant ribosomes that may be used as drug targets to identify drug candidates that are effective against the selected species. Examples of species include, without limitation, *Mycobacterium tuberculosis* (tuberculosis), *Pseudomonas aeruginosa* (multidrug resistant nosocomial infections), *Salmonella typhi* (typhoid fever), *Yersenia pestis* (plague), *Staphylococcus aureus* (multidrug resistant infections causing impetigo, folliculitis, abcesses, boils, infected lacerations, endocarditis, meningitis, septic arthritis, pneumonia, osteomyelitis, and toxic shock), *Streptococcus pyogenes* (streptococcal sore throat, scarlet fever, impetigo, erysipelas, puerperal fever, and necrotizing fascitis), *Enterococcus faecalis* (vancomycin resistant nosocomial infections, endocarditis, and bacteremia), *Chlamydia trachomatis* (lymphogranuloma venereum, trachoma and inclusion conjunctivitis, nongonococcal urethritis, epididymitis, cervicitis, urethritis, infant pneumonia, pelvic inflammatory diseases, Reiter's syndrome (oligoarthritis) and neonatal conjunctivitis), *Saccharomyces cerevesiae, Candida albicans,* and trypanosomes. In one embodiment, the rRNA gene is from *Mycobacterium tuberculosis* (see, *e.g.*, Example 6 and Figure 17).

In still other embodiments of the invention, the rRNA genes are mitochondrial rRNA genes, *i.e.*, eukaryotic rRNA genes (*e.g.*, human mitochondrial rRNA genes).

The plasmid constructs of the invention, such as the plasmid constructs set forth in Figures 22-26, may include novel mutant ASD and SD sequences set forth herein. In particular, the present invention provides novel mutant ASD sequences and novel mutant SD sequences, set forth in Figures 12, 13, 15, and 16, which may be used in the plasmid constructs and methods of the invention. The mutant ASD and mutant SD sequences

may be used as mutually compatible pairs (see Figures 12, 13, 15, and 16). It will be appreciated that the mutually compatible pairs of mutant ASD and SD sequences interact as pairs in the form of RNA and permit translation of only the mRNAs containing the compatible mutant SD sequence.

In another aspect, the present invention provides a plasmid comprising an *E. coli* 16S rRNA gene having a mutant ASD sequence, at least one mutation in said 16S rRNA gene, and a genetically engineered gene which encodes a selectable marker, *e.g.*, GFP, having a mutant SD sequence. In another embodiment, the 16S rRNA gene is from a species other than *E. coli*. In one embodiment, the mutant ASD sequence is selected from the sequences set forth in Figures 12, 13, 15, and 16. In another embodiment, the mutant SD sequence is selected from the sequences set forth in Figures 12, 13, 15, and 16. In yet another embodiment, the mutant ASD sequence and the mutant SD sequence are in mutually compatible pairs (see Figures 12, 13, 15, and 16). Each mutually compatible mutant SD and mutant ASD pair permits translation by the selectable marker.

In one embodiment, the invention features a cell comprising a plasmid of the invention. In another embodiment, the cell is a bacterial cell.

In one embodiment, the invention provides a method for identifying functional mutant ribosomes comprising:

(a) transforming a host cell with a plasmid comprising an rRNA gene having a mutant ASD sequence, at least one mutation in said rRNA gene, and a genetically engineered gene which encodes a selectable marker having a mutant SD sequence, wherein the mutant ASD and mutant SD sequences are a mutually compatible pair;

(b) isolating cells via the selectable marker; and

(c) identifying the rRNA from the cells from step (b), thereby identifying functional mutant ribosomes.

In another embodiment, the invention features a method for identifying functional mutant ribosomes comprising:

(a) transforming a host cell with a plasmid comprising an *E. coli* 16S rRNA gene having a mutant ASD sequence, at least one mutation in said 16S rRNA gene, and a genetically engineered gene which encodes GFP having a mutant SD sequence wherein the mutant ASD and mutant SD sequences are a mutually compatible pair;

(b) isolating cells via the GFP; and

(c) identifying the rRNA from the cells from step (b), thereby identifying functional mutant ribosomes.

In yet another embodiment, the invention features a method for identifying functional mutant ribosomes that may be suitable as drug targets comprising:

(a) transforming a host cell with a plasmid comprising an rRNA gene having a mutant ASD sequence, at least one mutation in said rRNA gene, and a genetically engineered gene which encodes a selectable marker having a mutant SD sequence, wherein the mutant ASD and mutant SD sequences are a mutually compatible pair;

(b) isolating cells via the selectable marker;

(c) identifying and sequencing the rRNA from the cells from step (b), thereby identifying regions of interest;

(d) selecting regions of interest from step (c);

(e) mutating the regions of interest from step (d);

(f) inserting the mutated regions of interest from step (e) into a plasmid comprising an rRNA gene having a mutant ASD sequence and a genetically engineered gene which encodes a selectable marker having a mutant SD sequence, wherein the mutant ASD and mutant SD sequences are a mutually compatible pair;

(g) transforming a host cell with the plasmid from step (f);

(h) isolating cells of step (g) via the selectable marker; and

(i) identifying the rRNA from step (h), thereby identifying functional mutant ribosomes that may be suitable as drug targets.

In a further embodiment, the invention provides a method for identifying functional mutant ribosomes that may be suitable as drug targets comprising:

(a) transforming a host cell with a plasmid comprising an *E. coli* 16S rRNA gene having a mutant ASD sequence, at least one mutation in said 16S rRNA gene, and a genetically engineered gene which encodes GFP having a mutant SD sequence wherein the mutant ASD and mutant SD sequences are a mutually compatible pair;

(b) isolating cells via the GFP;

(c) identifying and sequencing the rRNA from the cells from step (b), thereby identifying regions of interest;

(d) selecting the regions of interest from step (c);

(e) mutating the regions of interest from step (d);

(f) inserting the mutated regions of interest from step (e) into a plasmid comprising an *E. coli* 16S rRNA gene having a mutant ASD sequence and a genetically engineered gene which encodes GFP having a mutant SD sequence, wherein the mutant ASD and mutant SD sequences are a mutually compatible pair;

(g) transforming a host cell with the plasmid from step (f);

(h) isolating cells of step (g) via the GFP; and

(i) identifying the rRNA from step (h), thereby identifying functional mutant ribosomes that may be suitable as drug targets.

In one embodiment, the invention features a method for identifying drug candidates comprising:

(a) transforming a host cell with a plasmid comprising an rRNA gene having a mutant ASD sequence, at least one point mutation in said rRNA gene, and a genetically engineered gene which encodes a selectable marker having a mutant SD sequence, wherein the mutant ASD and mutant SD sequences are a mutually compatible pair;

(b) isolating cells via the selectable marker;

(c) identifying and sequencing the rRNA from step (b) to identify the regions of interest;

(d) selecting the regions of interest from step (c);

(e) mutating the regions of interest from step (d);

(f) inserting the mutated regions of interest from step (e) into a plasmid comprising an rRNA gene having a mutant ASD sequence and a genetically engineered gene which encodes a selectable marker having a mutant SD sequence, wherein the mutant ASD and mutant SD sequences are a mutually compatible pair;

(g) transforming a host cell with the plasmid from step (f);

(h) isolating the cells from step (g) via the selectable marker;

(i) identifying the rRNA from step (h) to identify the functional mutant ribosomes;

(j) screening drug candidates against functional mutant ribosomes from step (i);

(k) identifying the drug candidates from step (j) that bound to the functional mutant ribosomes from step (i);

(l) screening the drug candidates from step (k) against human rRNA; and

(m) identifying the drug candidates from step (l) that do not bind to human rRNA, thereby identifying drug candidates.

In one embodiment, the invention provides a method for identifying drug candidates comprising:

(a) transforming a host cell with a plasmid comprising an *E. coli* 16S rRNA gene having a mutant ASD sequence, at least one point mutation in said 16S rRNA gene, and a genetically engineered gene which encodes GFP having a mutant SD sequence, wherein the mutant ASD and mutant SD sequences are a mutually compatible pair;

(b) isolating the cells via the selectable marker;

(c) identifying and sequencing the rRNA from step (b) to identify the regions of interest;

(d) selecting the regions of interest from step (c);

(e) mutating the regions of interest from step (d);

(f) inserting the mutated regions of interest from step (e) into a plasmid comprising an *E. coli* 16S rRNA gene having a mutant ASD sequence and a genetically engineered gene which encodes GFP having a mutant SD sequence, wherein the mutant ASD and mutant SD sequences are a mutually compatible pair;

(g) transforming a host cell with the plasmid from step (f);

(h) isolating cells from step (g) via the selectable marker;

(i) identifying the rRNA from step (h) to identify the functional mutant ribosomes;

(j) screening drug candidates against the functional mutant ribosomes from step (i);

(k) identifying the drug candidates from step (j) that bound to the functional mutant ribosomes from step (i);

(l) screening the drug candidates from step (k) against human 16S rRNA; and

(m) identifying the drug candidates from step (l) that do not bind to the human 16S rRNA, thereby identifying drug candidates.

It will be appreciated that the rRNA gene used in the methods of the present invention may be from the 16S rRNA, 23S rRNA, and 55S rRNA gene.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 depicts the plasmid construct pRNA123. The locations of specific sites in pRNA123 are as follows: the 16S rRNA *E. coli* rrnB operon corresponds to nucleic acids 1-1542; the 16S MBS (message binding sequence) GGGAU corresponds to nucleic acids 1536-1540; the 16S-23S spacer region corresponds to nucleic acids 1543-1982; the 23S rRNA of *E. coli* rrnB operon corresponds to nucleic acids 1983-4886; the 23S-5S spacer region corresponds to nucleic acids 4887-4982; the 5S rRNA of *E. coli* rrnB operon corresponds to nucleic acids 4983-5098; the terminator T1 of *E. coli* rrnB operon corresponds to nucleic acids 5102-5145; the terminator T2 of *E. coli* rrnB operon corresponds to nucleic acids 5276-5305; the *bla* (β-lactamase; ampicillin resistance) corresponds to nucleic acids 6575-7432; the replication origin corresponds to nucleic acids 7575-8209; the *rop* (Rop protein) corresponds to nucleic acids 8813-8622; the GFP corresponds to nucleic acids 10201-9467; the GFP RBS (ribosome binding sequence) AUCCC corresponds to nucleic acids 10213-10209; the trp^c promoter corresponds to nucleic acids 10270-10230; the trp^c promoter corresponds to nucleic acids 10745-10785; the CAT RBS AUCCC corresponds to nucleic acids 10802-10806; the *cam* (chloramphenicol acetyltransferase: CAT) corresponds to nucleic acids 10814-11473; the $lacI^q$ promoter corresponds to nucleic acids 11782-11859; the $lacI^q$ (lac repressor) corresponds to nucleic acids 11860-12942; and the *lacUV5* promoter corresponds to nucleic acids 12985-13026.

Figure 2 depicts a scheme for construction of pRNA9. The abbreviations in Figure 2 are defined as follows: Ap^r, ampicillin resistance; *cam*, CAT gene; $lacI^q$, lactose repressor; P*lacUV5*, *lacUV5* promoter; $Ptrp^c$, constitutive *trp* promoter. The restriction sites used are also indicated.

Figure 3 depicts an autoradiogram of sequencing gels with pRNA8-rMBS-rRBS. The mutagenic MBS and RBS are shown: B 5 C, G, T; D 5 A, G, T; H 5 A, C, T; V 5 A, C, G. The start codon of *cam* and the 39 end of 16S rRNA are indicated. Panel A depicts the RBS of the CAT gene. Panel B depicts the MBS of the 16S rRNA gene.

Figure 4 depicts a graph of the effect of MBSs on growth. The abbreviations in Figure 4 are defined as follows: pBR322; vector: pRNA6; RBS 5 GUGUG, MBS 5 CACAC: pRNA9; RBS 5 GGAGG (wt), MBS 5 CCUCC (wt): and Clone IX24; RBS 5 AUCCC, MBS 5 GGGAU.

Figure 5 depicts a scheme for construction of pRNA122. The abbreviations in Figure 5 are defined as follows: Ap^r, ampicillin resistance; *cam*, CAT gene; $lacI^q$, lactose repressor; P*lacUV5*, *lacUV5* promoter; $Ptrp^c$, constitutive *trp* promoter; N 5 A, C, G, and T. The four nucleotides mutated are underlined and the restriction sites used are indicated.

Figure 6 depicts a plasmid-derived ribosome distribution and CAT activity. Cultures were induced (or not) in early log phase (as shown in Figure 4) and samples were withdrawn for CAT assay and total RNA preparation at the points indicated. Open squares represent the percent plasmid-derived rRNA in uninduced cells. Closed squares represent the percent plasmid-derived rRNA in induced cells. Open circles represent CAT activity in uninduced cells. Closed circles represent CAT activity in induced cells.

Figure 7 depicts a scheme for construction of single mutations at positions 516 or 535. The abbreviations in Figure 7 are defined as follows: Ap^r, ampicillin resistance; *cam*, CAT gene; $lacI^q$, lactose repressor; P*lacUV5*, *lacUV5* promoter; $Ptrp^c$, constitutive *trp* promoter. C516 was substituted to V (A, C, or G) and A535 was substituted to B (C, G, or T,) in pRNA122 and the restriction sites that were used are also indicated.

Figure 8 depicts the functional analysis of mutations constructed at positions 516 and 535 of 16S rRNA in pRNA122. Nucleotide identities are indicated in the order of 516:535 and mutations are underlined. pRNA122 containing the wild-type MBS (wt. MBS) was used as a negative control to assess the degree of MIC and the level of CAT activity due to CAT mRNA translation by wild-type ribosomes. Standard error of the mean is used to indicate the range of the assay results.

Figure 9 depicts a description and use of oligodeoxynucleotides. Primer binding sites are indicated by the number of nucleotides from the 5' nucleotide of the coding region. Negative numbers indicate binding sites 5' to the coding region.

Figure 10 describes several plasmids used in Example 4.

Figure 11 depicts the specificity of the selected recombinants. The concentrations of chloramphenicol used are indicated and the unit of MIC is micrograms of chloramphenicol/mL.

Figure 12 depicts novel mutant ASD sequences and novel mutant SD sequences of the present invention. Figure 12 also shows a sequence analysis of chloramphenicol resistant isolates. The mutated nucleotides are underlined and potential duplex formations are boxed. CAT activity was measured twice for each culture and the unit is CPM/0.1 μL of culture/OD600. Induction was measured by dividing CAT activity in induced cells with CAT activity in uninduced cells. A –1 indicates no induction, while a +1 indicates induction with 1 mM IPTG.

Figure 13 depicts novel mutant ASD sequences and novel mutant SD sequences of the present invention. Figure 13 also shows a sequence analysis of CAT mRNA mutants. Potential duplex formations are boxed and the mutated nucleotides are underlined. The start codon (AUG) is in bold. A –1 indicates no induction, while a +1 indicates induction with 1 mM IPTG.

Figure 14 depicts the effect of Pseudouridine516 Substitutions on subunit assembly. The percent plasmid-derived 30S data are presented as the percentage of the total 30S in each peak and in crude ribosomes.

Figure 15 depicts novel mutant ASD sequences and novel mutant SD sequences of the present invention.

Figure 16 depicts novel mutant ASD sequences and novel mutant SD sequences of the present invention.

Figure 17 depicts a hybrid construct. This hybrid construct contains a 16S rRNA from *Mycobacterium tuberculosis*. The specific sites on the hybrid construct are as follows: the part of rRNA from *E. coli* rrnB operon corresponds to nucleic acids 1-931; the part of 16S rRNA from *Mycobacterium tuberculosis* rrn operon corresponds to nucleic acids 932-1542; the 16S MBS (message binding sequence) GGGAU corresponds to nucleic acids 1536-1540; the terminator T1 of *E. coli* rrnB operon corresponds to nucleic acids 1791-1834; the terminator T2 of *E. coli* rrnB operon corresponds to nucleic acids 1965-1994; the replication origin corresponds to nucleic acids 3054-2438; the *bla* (β-lactamase; ampicillin resistance) corresponds to nucleic acids 3214-4074; the GFP corresponds to nucleic acids 5726-4992; the GFP RBS (ribosome binding sequence) AUCCC corresponds to nucleic acids 5738-5734; the trp^c promoter corresponds to nucleic acids 5795-5755; the trp^c promoter corresponds to nucleic acids 6270-6310; the CAT RBS (ribosome binding sequence) AUCCC corresponds to nucleic acids 6327-6331; the *cam* (chloramphenicol acetyltransferase; CAT) corresponds to nucleic acids 6339-6998; the $lacI^q$ promoter corresponds to nucleic acids 7307-7384; the $lacI^q$ (lac

repressor) corresponds to nucleic acids 7385-8467; and the *lacUV5* promoter corresponds to nucleic acids 8510-8551.

Figure 18 depicts a plasmid map of pRNA122.

Figure 19 depicts a table of sequences and MICs of functional mutants. Sequences are ranked by the minimum inhibitory concentration ("MIC") of chloramphenicol required to fully inhibit growth of cells expressing the mutant ribosomes. The nucleotide sequences ("Nucleotide sequence") are the 790 loop sequences selected from the pool of functional, randomized mutants. Mutations are underlined. The number of mutations ("Number of mutations") in each mutant sequence are indicated, as well as the number of occurrences ("Number of occurrences") which represents the number of clones with the indicated sequence. The sequence and activity of the unmutated control, pRNA122 (WT, wild-type) is depicted in the first row of Figure 19, in which the MIC is 600 μg/ml.

Figure 20 depicts the 790-loop sequence variation. In the consensus sequence R = A or G; N = A, C, G or U; M = A or C; H = A, C or U; W = A or U; Y = C or U; Δ= deletion; and underlined numbers indicate the wild-type *E. coli* sequence.

Figure 21 depicts functional and thermodynamic analysis of positions 787 and 795. Mutations have been underlined and "n.d." represents not determined. Figure 21 shows site-directed mutations ("Nucleotide") that were constructed using PCR, as described for the random mutants, except that the mutagenic primers contained substitutions corresponding only to positions 787 and 795. In order to determine ribosome function ("Mean CAT activity"), each strain was grown and assayed for CAT activity at least twice, the data were averaged, and presented as percentages of the unmutated control, pRNA122 ± the standard error of the mean. The ratio of plasmid to chromosome-derived rRNA in 30S and 70 S ribosomes ("% Mutant 30S in 30S peak/ 70S peak") was determined by primer extension. Cultures were grown and assayed at least twice and the mean values are presented as a percentage of the total 30S in each peak ± the standard error of the mean. Thermodynamic parameters ("Thermodynamics") are for the higher-temperature transition of model oligonucleotides and are the average of results for four or five different oligomer concentrations. Standard errors for the $\Delta G°37$ are ± 5% (1 kcal = 4184 J). Errors in T_m are estimated as ± 1 °C. All solutions were at pH 7.

Figure 22 depicts the DNA sequence of pRNA8.

Figure 23 depicts the DNA sequence of pRNA122.

Figure 24 depicts the DNA sequence of pRNA123.

Figure 25 depicts the DNA sequence of pRNA123 *Mycobacterium tuberculosis* - 2 (pRNA123 containing a hybrid of *E. coli* and *Mycobacterium tuberculosis* 16S rRNA genes).

Figure 26 depicts the DNA sequence of pRep- *Mycobacterium tuberculosis-2* (containing a puc19 derivative containing the rRNA operon from pRNA122; however, the 23S and 5S rRNA genes are deleted).

Figures 2-14 may be found in Lee, K., *et al.* Genetic Approaches to Studying Protein Synthesis: Effects of Mutations at Pseudouridine 516 and A535 in *Escherichia coli* 16S rRNA. *Symposium: Translational Control: A Mechanistic Perspective* at the Experimental Biology 2001 Meeting (2001); and Figures 18-21 may be found in Lee, K. *et al.*, *J. Mol. Biol.* 269: 732-743 (1997), all of which are expressly incorporated by reference herein.

Detailed Description of the Invention

Compositions and methods are provided to identify functional mutant ribosomes suitable as drug targets. The compositions and methods allow isolation and analysis of mutations that would normally be lethal and allow direct selection of rRNA mutants with predetermined levels of ribosome function. The compositions and methods of the present invention may be used to identify antibiotics to treat generally and/or selectively human pathogens.

According to one embodiment of the invention, a functional genomics database for rRNA genes of a variety of species may be generated. In particular, the rRNA gene is randomly mutated using a generalized mutational strategy. A host cell is then transformed with a mutagenized plasmid of the invention comprising: an rRNA gene having a mutant ASD sequence, the mutated rRNA gene, and a genetically engineered gene which encodes a selectable marker having a mutant SD sequence. The selectable marker gene, such as CAT, may be used to select mutants that are functional, *e.g.*, by plating the transformed cells onto growth medium containing chloramphenicol. The mutant rRNA genes contained in each plasmid DNA of the individual clones from each colony are selected and characterized. The function of each of the mutant rRNA genes is assessed by measuring the amount of an additional selectable marker gene, such as GFP, produced by each clone upon induction of the rRNA operon. A functional genomics database may thus be assembled, which contains the sequence and functional data of the functional mutant rRNA genes. In particular, functionally important regions of the rRNA gene that will serve as drug targets are identified by comparing the sequences of the functional genomics database and correlating the sequence with the amount of GFP protein produced.

In another embodiment, the nucleotides in the functionally important target regions identified in the above methods may be simultaneously randomly mutated, *e.g.*, by using standard methods of molecular mutagenesis, and cloned into a plasmid of the invention to form a plasmid pool containing random mutations at each of the nucleotide

positions in the target region. The resulting pool of plasmids containing random mutations is then used to transform cells, *e.g.*, *E. coli* cells, and form a library of clones, each of which contains a unique combination of mutations in the target region. The library of mutant clones are grown in the presence of IPTG to induce production of the mutant rRNA genes and a selectable marker is used, such as CAT, to select clones of rRNA mutants containing nucleotide combinations of the target region that produce functional ribosomes. The rRNA genes producing functional ribosomes are sequenced and may be incorporated into a database.

In yet another embodiment, a series of oligonucleotides may be synthesized that contain the functionally-important nucleotides and nucleotide motifs within the target region and may be used to sequentially screen compounds and compound libraries to identify compounds that recognize (bind to) the functionally important sequences and motifs. The compounds that bind to all of the oligonucleotides are then counterscreened against oligonucleotides and/or other RNA containing molecules to identify drug candidates. Drug candidates selected by the methods of the present invention are thus capable of recognizing all of the functional variants of the target sequence, *i.e.*, the target cannot be mutated in a way that the drug cannot bind, without causing loss of function to the ribosome.

In still another embodiment, after the first stage mutagenesis of the entire rRNA is performed using techniques known in the art, *e.g.*, error-prone PCR mutagenesis, the mutants are analyzed to identify regions within the rRNA that are important for function. These regions are then sorted based on their phylogenetic conservation, as described herein, and are then used for further mutagenesis.

Ribosomal RNA sequences from each species are different and the more closely related two species are, the more their rRNAs are alike. For instance, humans and monkeys have very similar rRNA sequences, but humans and bacteria have very different rRNA sequences. These differences may be utilized for the development of very specific drugs with a narrow spectrum of action and also for the development of broad-spectrum drugs that inhibit large groups of organisms that are only distantly related, such as all bacteria.

In another embodiment, the functionally important regions identified above are divided into groups based upon whether or not they occur in closely related groups of organisms. For instance, some regions of rRNA are found in all bacteria but not in other organisms. Other areas of rRNA are found only in closely related groups of bacteria, such as all of the members of a particular species, *e.g.*, members of the genus *Mycobacterium* or *Streptococcus*.

In a further embodiment, the regions found in very large groups of organisms, *e.g.*, all bacteria or all fungi, are used to develop broad-spectrum antibiotics that may be

used to treat infections from a large number of organisms within that group. The methods of the present invention may be performed on these regions and functional mutant ribosomes identified. These functional mutant ribosomes may be screened, for example, with compound libraries.

In yet another embodiment, regions that are located only in relatively small groups of organisms, such as all members of the genus *Streptococcus* or all members of the genus *Mycobacterium,* may be used to design narrow spectrum antibiotics that will only inhibit the growth of organisms that fall within these smaller groups. The methods of the present invention may be performed on these regions and functional mutant ribosomes identified. These functional mutant ribosomes will be screened, *e.g.*, compound libraries.

The invention provides novel plasmid constructs, *e.g.* pRNA123 (Figures 1 and 24). The novel plasmid constructs of the present invention employ novel mutant ASD and mutant SD sequences set forth in Figures 12, 13, 15 and 16. The mutant ASD and mutant SD sequences may be used as mutually compatible pairs (see Figures 12, 13, 15 and 16). It will be appreciated that the mutually compatible pairs of mutant ASD and SD sequences interact as pairs in the form of RNA, to permit translation of only the mRNAs containing the altered SD sequence.

Definitions

As used herein, each of the following terms has the meaning associated with it in this section.

The articles "a" and "an" are used herein to refer to one or to more than one (*i.e.* to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

An "inducible" promoter is a nucleotide sequence which, when operably linked with a polynucleotide which encodes or specifies a gene product, causes the gene product to be produced in a living cell substantially only when an inducer which corresponds to the promoter is present in the cell.

As used herein, the term "mutation" includes an alteration in the nucleotide sequence of a given gene or regulatory sequence from the naturally occurring or normal nucleotide sequence. A mutation may be a single nucleotide alteration (*e.g.*, deletion, insertion, substitution, including a point mutation), or a deletion, insertion, or substitution of a number of nucleotides.

By the term "selectable marker" is meant a gene whose expression allows one to identify functional mutant ribosomes.

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term "isolated nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium, when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence set forth in Figures 12, 13, 15, and 16, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or portion of the nucleic acid sequence set forth in Figures 12, 13, 15, and 16 as a hybridization probe, the nucleic acid molecules of the present invention can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of the sequence set forth in Figures 12, 13, 15, and 16 can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence set forth in Figures 12, 13, 15, and 16.

A nucleic acid of the invention can be amplified using cDNA, mRNA or, alternatively, genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to the nucleotide sequences of the present invention can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence set forth in Figures 12, 13, 15, and 16, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in Figures 12, 13, 15, and 16, is one which is sufficiently complementary to the nucleotide sequence shown in Figures 12, 13, 15, and 16, such that it can hybridize to the nucleotide sequence shown in Figures 12, 13, 15, and 16, respectively, thereby forming a stable duplex.

II. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid molecule of the present invention (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a

host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel (1990) *Methods Enzymol.* 185:3-7. Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.* (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.* (1990) *Methods Enzymol.* 185:60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the

recombinant protein (Gottesman, S. (1990) *Methods Enzymol.* 185:119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.* (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the expression vector may be a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari *et al.* (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.* (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (Invitrogen Corp, San Diego, CA).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J. *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example by the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379).

Another aspect of the invention pertains to host cells into which a the nucleic acid molecule of the invention is introduced. The terms "host cell" and "recombinant

host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

III. Uses and Methods of the Invention

The nucleic acid molecules described herein may be used in a plasmid construct, *e.g.* pRNA123, to carry out one or more of the following methods: (1) creation of a functional genomics database of the rRNA genes generated by the methods of the present invention; (2) mining of the database to identify functionally important regions of the rRNA; (3) identification of functionally important sequences and structural motifs within each target region; (4) screening compounds and compound libraries against a series of functional variants of the target sequence to identify compounds that bind to all functional variants of the target sequence; and (5) counterscreening the compounds against nontarget RNAs, such as human ribosomes or ribosomal RNA sequences.

This invention is further illustrated by the following examples, which should not be construed as limiting. The contents of all references, patents and published patent

applications cited throughout this application, as well as the Figures and Appendices, are incorporated herein by reference.

SPECIFIC EXAMPLES

EXAMPLE 1: IDENTIFICATION OF MUTANT SD AND MUTANT ASD COMBINATIONS

It has been shown that by coordinately changing the SD and ASD, a particular mRNA containing an altered SD could be targeted to ribosomes containing the altered ASD. This and all other efforts to modify the ASD, however, have proved lethal, as cells containing these mutations died within two hours after the genes containing them were activated.

Using random mutagenesis and genetic selection, mutant SD-ASD combinations were screened in order to identify nonlethal SD-ASD combinations. The mutant SD-ASD mutually compatible pairs are set forth in Figures 12, 13 15 and 16. The mutually compatible pairs of mutant sequences interact as pairs in the form of RNA. The novel mutant SD-ASD sequence combinations of the present invention permit translation of only the mRNAs containing the altered SD sequence.

EXAMPLE 2: CONSTRUCTION OF THE pRNA123 PLASMID

A plasmid construct of the present invention identified as the pRNA123 plasmid, is set forth in Figures 1 and 24. *E. coli* cells contain a single chromosome with seven copies of the rRNA genes and all of the genes for the ribosomal proteins. The plasmid, pRNA123, in the cell contains a genetically engineered copy of one of the rRNA genes from *E. coli* and two genetically engineered genes that are not normally found in *E. coli,* referred to herein as a "selectable markers." One gene encodes the protein chloramphenicol acetyltransferase (CAT). This protein renders cells resistant to chloramphenicol by chemically modifying the antibiotic. Another gene, the Green Fluorescent Protein (GFP), is also included in the system. GFP facilitates high-throughput functional analysis. The amount of green light produced upon irradiation with ultraviolet light is proportional to the amount of GFP present in the cell.

Ribosomes from pRNA123 have an altered ASD sequence. Therefore, the ribosomes can only translate mRNAs that have an altered SD sequence. Only two genes in the cell produce mRNAs with altered SD sequences that may be translated by the plasmid-encoded ribosomes: the CAT and GFP gene. Mutations in rRNA affect the ability of the resulting mutant ribosome to make protein. The present invention thus

provides a system whereby the mutations in the plasmid-encoded rRNA gene only affect the amount of GFP and CAT produced. A decrease in plasmid ribosome function makes the cell more sensitive to chloramphenicol and reduces the amount of green fluorescence of the cells. Translation of the other mRNAs in the cell is unaffected since these mRNAs are translated only by ribosomes that come from the chromosome. Hence, cells containing functional mutants may be identified and isolated via the selectable marker.

EXAMPLE 3: GENETIC SYSTEM FOR FUNCTIONAL ANALYSIS OF RIBOSOMAL RNA

Identification of Functionally Important Regions of rRNA. Functionally important regions of rRNA molecules that may be used as drug targets using a functional genomics approach may be identified through a series of steps. Namely, in step I.a., the entire rRNA gene is randomly mutated using error-prone PCR or another generalized mutational strategy. In step I.b., a host cell is then transformed with a mutagenized plasmid comprising: an rRNA gene having a mutant ASD sequence, at least one mutation in said rRNA gene, and a genetically engineered gene which encodes a selectable marker having a mutant SD sequence, and production of the rRNA genes from the plasmid are induced by growing the cells in the presence of IPTG. In step I.c., the CAT gene is used to select mutants that are functional by plating the transformed cells onto growth medium containing chloramphenicol. In step I.d., individual clones from each of the colonies obtained in step I.c. are isolated. In step I.e., the plasmid DNA from each of the individual clones from step I.d. is isolated. In step I.f., the rRNA genes contained in each of the plasmids that had been isolated in step I.e. are sequenced. In step I.g., the function of each of the mutants from step I.f. is assessed by measuring the amount of GFP produced by each clone from step I.e. upon induction of the rRNA operon. In step I.h., a functional genomics database is assembled containing the sequence and functional data from steps I.f. and I.g. In step I.i., functionally important regions of the rRNA gene that will serve as drug targets are identified. Functionally important regions may be identified by comparing the sequences of all of the functional genomics database constructed in step I.g. and correlating the sequence with the amount of GFP protein produced. Contiguous sequences of three or more rRNA nucleotides, in which substitution of the nucleotides in the region produces significant loss of function, will constitute a functionally important region and therefore a potential drug target.

Isolation of Functional Variants of the Target Regions. A second aspect of the invention features identification of mutations of the target site that might lead to antibiotic resistance using a process termed, "instant evolution", as described below. In

step II.a., for a given target region identified in step I.i., each of the nucleotides in the target region is simultaneously randomly mutated using standard methods of molecular mutagenesis, such as cassette mutagenesis or PCR mutagenesis, and cloned into the plasmid of step I.b. to form a plasmid pool containing random mutations at each of the nucleotide positions in the target region. In step II.b., the resulting pool of plasmids containing random mutations from step II.a. is used to transform *E. coli* cells and form a library of clones, each of which contains a unique combination of mutations in the target region. In step II.c., the library of mutant clones from step II.b. is grown in the presence of IPTG to induce production of the mutant rRNA genes. In step II.d., the induced mutants are plated on medium containing chloramphenicol, and CAT is used to select clones of rRNA mutants containing nucleotide combinations of the target region that produce functional ribosomes. In step II.e., the functional clones isolated in step II.d. are sequenced and GFP is used to measure ribosome function in each one. In step II.f., the data from step II.e. are incorporated into a mutational database.

Isolation of Drug Leads. In step III.a., the database in step II.f. is analyzed to identify functionally-important nucleotides and nucleotide motifs within the target region. In step III.b., the information from step III.a. is used to synthesize a series of oligonucleotides that contain the functionally important nucleotides and nucleotide motifs identified in step III.a. In step III.c., the oligonucleotides from step III.b. are used to sequentially screen compounds and compound libraries to identify compounds that recognize (bind to) the functionally important sequences and motifs. In step III.d., compounds that bind to all of the oligonucleotides are counterscreened against oligonucleotides and/or other RNA containing molecules to identify drug candidates. "Drug candidates" are compounds that 1) bind to all of the oligonucleotides containing the functionally important nucleotides and nucleotide motifs, but do not bind to molecules that do not contain the functionally important nucleotides and nucleotide motifs and 2) do not recognize human ribosomes. Drug candidates selected by the methods of the present invention therefore recognize all of the functional variants of the target sequence, *i.e.*, the target cannot be mutated in a way that the drug cannot bind, without causing loss of function to the ribosome.

EXAMPLE 4: GENETIC SYSTEM FOR STUDYING PROTEIN SYNTHESIS

Materials and Methods

Reagents. All reagents and chemicals were as in Lee, K., *et al.* (1996) *RNA* 2: 1270–1285. PCR-directed mutagenesis was performed essentially by the method of Higuchi, R. (1989) *PCR Technology* (Erlich, H. A., ed.), pp. 61–70. Stockton Press, New

York, NY. The primers used in the present invention are listed in Figure 9. The plasmids used in the present invention are listed in Figure 10.

Bacterial strains and media. All plasmids were maintained and expressed in *E. coli* DH5 (*supE44, hsdR17, recA1, endA1, gyrA96, thi-1* and *relA1*) (36). To induce synthesis of plasmid-derived rRNA from the *lacUV5* promoter, IPTG was added to a final concentration of 1 mM. Chloramphenicol acetyltransferase activity was determined essentially as described by Nielsen *et al.* (1989) *Anal. Biochem.* 179: 19–23. Cultures for CAT assays were grown in LB-Ap100. MIC were determined by standard methods in microtiter plates as described in Lee, K., *et al.* (1997) *J. Mol. Biol.* 269: 732–743.

Primer extension. To determine the ratio of plasmid to chromosome-derived rRNA, pRNA104 containing cells growing in LB-Ap100 were harvested at the time intervals indicated and total RNA was extracted using the Qiagen RNeasy kit (Chatsworth, CA). The 30S, 70S, and crude ribosomes were isolated from 200 mL of induced, plasmid containing cells by the method of Powers and Noller (Powers, T. *et al.* (1991) *EMBO J.* 10: 2203–2214). The purified RNA was analyzed by primer extension according to Sigmund, C. D., *et al.* (1988) *Methods Enzymol.* 164: 673–690.

Experimental Procedures

Generation of pRNA9 construct. The initial construct, pRNA9, was generated using the following methods. Plasmid pRNA9 contains a copy of the *rrnB* operon from pKK3535 under transcriptional regulation of the *lacUV5* promoter; this well-characterized promoter is not subject to catabolic repression and is easily and reproducibly inducible with isopropyl-β-D-thiogalactoside (IPTG). To minimize transcription in the absence of inducer, PCR was used to amplify and subclone the *lac* repressor variant, $lacI^q$ (Calos, M. P. (1978) *Nature* 274: 762–765) from pSPORT1 (Life Technologies, Rockville, MD). The chloramphenicol acetyltransferase gene (*cam*) is present and transcribed constitutively from a mutant tryptophan promoter, trp^c (De Boer, H. A., *et al.* (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80: 21–25; Hui, A., *et al.* (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84: 4762–4766). The β-lactamase gene is also present to allow maintenance of plasmids in the host strain. To allow genetic selection, the CAT structural gene from pJLS1021 (Schottel, J. L., *et al.* (1984) *Gene* 28: 177–193) was amplified and placed downstream of a constitutive trp^c promoter using PCR. Expression of the CAT gene in *E. coli* renders the cell resistant to chloramphenicol and the minimal inhibitory concentration, hereinafter referred to as MIC, of chloramphenicol increases proportionally with the amount of CAT protein produced (Lee, K., *et al.* (1996) *RNA* 2: 1270–1285; Lee, K., *et al.* (1997) *J. Mol. Biol.* 269: 732–743) An overview of the steps used to construct the system is shown in Figure 2.

Selection of a new MBS-RBS pair. To isolate message binding site-ribosome binding site, hereinafter referred to as MBS-RBS, combinations that are nonlethal and efficiently translated only by plasmid-derived ribosomes, a random mutagenesis and selection scheme were used. In particular, the plasmid-encoded 16S MBS and CAT RBS were randomly mutated using PCR so that the wild-type nucleotide at each position was excluded. An autoradiogram of sequencing gels with pRNA8-rMBS-rRBS is provided in Figure 3. The resulting 2.5 x10^6 doubly mutated transformants were induced for 3.5 hours in SOC medium containing 1 mM IPTG and plated on Luria broth medium containing 100 μg/mL ampicillin, 350 μg/mL chloramphenicol and 1 mM IPTG. To confirm the presence of all three alternative nucleotides at each mutated position, plasmid DNA from approximately 2.0 x 10^5 transformants was sequenced (Figure 3).

Results

The data show that all of the nonexcluded nucleotides were equally represented in the random pool. Of the 2.5 x10^6 transformants plated, 536 survived the chloramphenicol selection. The efficiency of the selected MBS-RBS combinations was determined by measuring the minimal inhibitory concentration, hereinafter referred to as MIC, of chloramphenicol for each survivor in the presence and absence of inducer (Figure 11) (Lee, K., *et al.* (1996) *RNA* 2: 1270–1285; Lee, K., *et al.* (1997) *J. Mol. Biol.* 269: 732–743). Nine of the isolates (1.7%) showed MIC in the presence of inducer, which were lower than the 350 μg/mL concentration at which they were selected. These were slow growing mutants that appeared after 48 hours during the initial isolation. The MIC, however, were scored after only 24 hours. The MIC for 451 of the isolates (84.1%) were between 400 and 600 μg/mL, and the remaining 76 clones (14.2%) were 600 μg/mL. The difference in chloramphenicol resistance between induced and uninduced cells (ΔMIC) is the amount of CAT translation by plasmid-derived ribosomes only. A specific interaction between plasmid-derived ribosomes and CAT mRNA was indicated in 79 (14.7%) of the clones, which showed four- to eightfold increases in CAT resistance upon addition of IPTG (Figure 11).

Based on these analyses, 11 clones were retained for additional study. The MBS and RBS in plasmids from these clones were sequenced and CAT assays and growth curves were performed (Figures 4 and 12). Although a wide range of inducibility was observed, there was no correlation between specificity and predicted free energy ($\Delta G^{\circ 37}$). Purines were preferred in all of the MBS positions, but the RBS did not show this sort of selectivity. This can be explained partially by the observation that the selected RBS can base pair with sequences adjacent to the mutated region of 16S rRNA (Lee, K., *et al.* (1996) *RNA* 2: 1270–1285).

Growth curves were performed for all of the selected mutants and compared with strains containing control constructs (Figure 4). Only one mutant (IX24) is shown in Figure 4, but all strains containing the selected MBS/RBS sequences showed the same pattern of growth as this mutant. Because of its induction profile, strain IX24 (containing plasmid pRNA100) was chosen for additional experimentation. To eliminate the possibility that mutations outside the MBS and RBS had been inadvertently selected, the *Dra*III and *Xba*I fragment containing the MBS and the *Kpn*I and *Xho*I fragment containing the RBS sequence from pRNA100 (Figure 5) were transferred to pRNA9.

Specificity of the system. The rate of ribosome induction and the ratio of plasmid to chromosome-derived rRNA at each stage of growth were determined. For this, a pRNA100 derivative, pRNA104, which contains a C1192U mutation in 16S rRNA was constructed (Sigmund, C. D., *et al.* (1984) *Nucleic Acids Res.* 12: 4653–4663; Triman, K., *et al.* (1989) *J. Mol. Biol.* 209: 645–653) so that plasmid-derived rRNA could be differentiated from wild-type rRNA by primer extension. The C1192U mutation does not affect ribosome function in other expression systems (Sigmund, C. D., *et al.* (1984) *Nucleic Acids Res.* 12: 4653–4663; Makosky, P. C. *et al.* (1987) *Biochimie* 69: 885–889). To show that the same is true in the present system, CAT activity was measured after 3 hours induction with 1 mM IPTG in DH5 cells expressing pRNA100 or pRNA104 and the two were compared. In these experiments, no significant difference between cells expressing pRNA104 (99.2 $\pm$ 2.8%) or pRNA100 (100%) was observed.

To determine the percentage of plasmid-derived ribosomes in cells containing the plasmid, total RNA was isolated from DH5 cells carrying pRNA104 before and after induction with IPTG and subjected to primer extension analysis (Lee, K., *et al.* (1997) *J. Mol. Biol.* 269: 732–743; Sigmund, C. D., *et al.* (1984) *Nucleic Acids Res.* 12: 4653–4663; Makosky, P. C. *et al.* (1987) *Biochimie* 69: 885–889). Maximum induction of plasmid-derived ribosomes occurred 3 hours after induction at which point they constituted approximately 40% of the total ribosome pool (Figure 6). CAT activities in these cells paralleled induction of plasmid-derived ribosomes and began to decrease 4 hours after induction, presumably due to protein degradation during stationary phase. In uninduced cells, approximately 3% of the total ribosome pool contains plasmid-derived ribosomes because of basal level transcription from the *lacUV5* promoter.

Optimization of the system. Chloramphenicol resistance in uninduced cells containing pRNA100 is 75 μg/mL (Figure 13, MIC = 100 μg/mL). By measuring CAT resistance in a derivative of pRNA100 containing a wild-type 16S rRNA gene, it was determined that approximately one-half of this background activity was due to CAT

translation by wild-type ribosomes (Figure 13, pRNA100 1 wt MBS). The remaining activity in uninduced cells is presumably due to leakiness of the *lacUV5* promoter (Figure 6). The nucleotide sequence located between the RBS and the start codon in mRNA affects translational efficiency (Calos, M. P. (1978) *Nature* 274: 762–765; Stormo, G. D., *et al.* (1982) *Nucleic Acids Res.* 10: 2971–2996; Chen, H., *et al.* (1994) *Nucl. Acids Res.* 22: 4953–4957). In pRNA100, three of the nucleotides found in this region of the CAT mRNA are complementary with the 3' terminus of wild-type *E. coli* 16S RNA (Figure 11, pRNA100 1 wt MBS). To eliminate the possibility that this was contributing to CAT translation in the absence of plasmid-encoded ribosomes, four nucleotides in the CAT gene (underlined in Figure 11) were randomly mutagenized and screened to identify mutants with reduced translation by host ribosomes. A total of 2000 clones were screened in the absence of plasmid-encoded ribosomes using pCAM9 and six poorly translated CAT sequences were isolated (Figure 13). Next, the *Bam*HI fragment of pRNA100 containing $lacI^q$ and the *rrnB* operon was added, and MIC, CAT assays and growth curves were performed on cells expressing these constructs (data not shown).

Based on these data, pRNA122 was chosen because it produced a slightly better induction profile than the others (Figures 11 and 23). Translation of the pRNA122 CAT message by wild-type ribosomes (Figure 11, pRNA122 1 wt MBS) produces cells that are sensitive to chloramphenicol concentrations <10 μg/mL. In the presence of specialized ribosomes (Figure 13, pRNA122), the background chloramphenicol MIC is between 40 and 50 μg/mL and the MIC for induced cells is between 550 and 600 μg/mL, producing an approximately13-fold increase in CAT expression upon induction in pRNA122. Induction of the *rrnB* operon in pRNA100 produces only an eightfold increase.

Use of the system. To test the system, the effects of nucleotide substitutions at the sole pseudouridine in *E. coli* 16S rRNA, located at position 516 were examined. Because pseudouridine and U form equally stable base pairs with adenosine (Maden, B. E. (1990) *Prog. Nucleic Acid Res.* Mol. Biol. 39: 241–303), mutations at A535 were also constructed to determine whether the potential for base pair formation between these two loci affected ribosome function. The mutations were constructed initially in a pUC19 (Yanisch-Perron, C., *et al.* (1985) *Gene* 33: 103–119) derivative containing the 16S RNA gene, p16ST, as shown in Figure 7 and then transferred to pRNA122 for analysis. This two-step process was used, because the *Sac*II restriction site located between the two mutated positions is unique in pRNA16ST and is not unique in pRNA122. The effect of the mutations in pRNA122 on protein synthesis *in vivo* was determined by measuring the MIC and CAT activity of the mutant cells (Figure 8). At position 516,

ribosomes containing the single transition mutation, pseudouridine516C, produced approximately 60% of the amount of functional CAT protein produced by wild-type ribosomes. The transversion mutations, pseudouridine516A or pseudouridine516G, however, reduced ribosome function by> 90%. All of the single mutations at position 535 retained > 50% of the function of wild-type ribosomes. To examine the possibility that the potential for base pairing between positions 516 and 535 is necessary for ribosome function, all possible mutations between these loci were also constructed and analyzed (Figure 8). These data show that all of the double mutants were inactive (10% or less of the wild-type) regardless of the potential to base pair. To examine the reasons for loss of function in the 516 mutants, ribosomes from cells expressing single mutations at position 516 were fractionated by sucrose density gradient centrifugation and the 30S and 70S peaks were analyzed by primer extension to determine the percentage of plasmid-derived 30S subunits present. The data in Figure 14 show a strong correlation between ribosome function and the presence of plasmid-derived ribosomes in the 70S ribosomal fraction, indicating that mutations at positions 516 affect the ability of the mutant 30S subunits to form 70S ribosomes.

The references cited in Example 4 may be found in Lee, K., *et al.* Genetic Approaches to Studying Protein Synthesis: Effects of Mutations at Pseudouridine516 and A535 in *Escherichia coli* 16S rRNA. *Symposium: Translational Control: A Mechanistic Perspective* at the Experimental Biology 2001 Meeting (2001) and at Lee, K. *et al.* (2001) Genetic Approaches to Studying Protein Synthesis: Effects of Mutations at Pseudouridine516 and A535 in *Escherichia coli* 16S rRNA. *J. Nutrition* 131 (11):2994-3004.

EXAMPLE 5: *IN VIVO* DETERMINATION OF RNA STRUCTURE-FUNCTION RELATIONSHIPS

Materials and Methods

Reagents. Restriction enzymes, ligase, AMV reverse transcriptase and calf intestine alkaline phosphatase were from New England Biolabs and from Gibco-BRL. Sequenase modified DNA polymerase, nucleotides and sequencing buffers were from USB/Amersham. Oligonucleotides were synthesized on-site using a Beckman Oligo 1000 DNA synthesizer. Amplitaq DNA polymerase and PCR reagents were from Perkin-Elmer-Cetus. [^{3}H]Chloramphenicol (30.1 Ci/mmol) was from Amersham and [α-^{35}S]dATP (1000 Ci/mmol) was from New England Nuclear. Other chemicals were from Sigma.

pRNA122. The key features of this construct are: (1) it contains a copy of the rrnB operon from pKK3535 (Brosius. J., *et al.* (1981) *Plasmid* 6:112-118.) under transcriptional regulation of the *lacUV5* promoter; (2) it contains a copy of the lactose repressor allele $lacI^q$ (Calos, M.P. (1978) *Nature* 274:762-769; (3) the chloramphenicol acetyltransferase gene (*cam*) is present and transcribed constitutively from a mutant tryptophan promoter, trp^c (de Boer, H. A., *et al.* (1983) *Proc. Natl Acad. Sci. USA* 80:21-25); (4) the RBS of the CAT message has been changed from the wild-type, 5' -GGAGG to 5' -AUCCC, and the MBS of the 16S rRNA gene has been changed to 5' -GGGAU; and (5) the β-lactamase gene is present to allow maintenance of plasmids in the host strain.

Bacterial strains and media. Plasmids were maintained and expressed in *E. coli* DH5 (*supE44, hsdR17, recA1, endA1, gyrA96, thi-1*; Hanahan, D. (1983) *J. Mol. Biol.* 166:557-580). Cultures were grown in LB medium (Luria, S.E. & Burrous, J.W. (1957) *J. Bacteriol.* 74:461-476) or LB medium containing 100 μg/ml ampicillin (LB-Ap100). To induce synthesis of plasmid-derived rRNA from the *lacUV5* promoter, IPTG was added to a final concentration of 1 mM at the times indicated in each experiment. Strains were transformed by electroporation (Dower, W. J., *et al.* (1988) *Nucl. Acids Res.* 16: 6127) using a Gibco-BRL Cell Porator. Unless otherwise indicated, transformants were grown in SOC medium (Hanahan, 1983, *supra*) for one hour prior to plating on selective medium to allow expression of plasmid-derived genes.

Chloramphenicol acetyltransferase assays. CAT activity was determined essentially as described (Nielsen, D. A. *et al.* (1989) *Anal. Biochem.* 60:191-227). Cultures for CAT assays were grown in LB-Ap100. Briefly, 0.5 ml aliquots of mid-log cultures (unless otherwise indicated) were added to an equal volume of 500 mM Tris-HCl (pH8) and lysed using 0.01% (w/v) SDS and chloroform (Miller, J.H. (1992) A Short Course in Bacterial Genetics, (Miller, J. H., ed.), pp. 71-80, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). The resulting lysate was either used directly or diluted in assay buffer prior to use. Assay mixtures contained cell extract (5 μl or 10 μl), 250 mM Tris (pH 8), 214 μM butyryl-coenzyme A (Bu-CoA), and 40 μM [^{3}H]chloramphenicol in a 125 μl volume. Two concentrations of lysate were assayed for one hour at 37°C to ensure that the signal was proportional to protein concentrations. The product, butyryl-[^{3}H]chloramphenicol was extracted into 2,6,10,14-tetramethylpentadecane:xylenes (2:1) and measured directly in a Beckman LS-3801 liquid scintillation counter. Blanks were prepared exactly as described above, except that uninoculated LB medium was used instead of culture.

Minimum inhibitory concentration determination. MICs were determined by standard methods in microtiter plates or on solid medium. Overnight cultures grown in LB-Ap100 were diluted and induced in the same medium containing 1 mM IPTG for three hours. Approximately 10^4 induced cells were then added to wells (or spotted onto solid medium) containing LB-Ap100 + IPTG (1 mM) and chloramphenicol at increasing concentrations. Cultures were grown for 24 hours and the lowest concentration of chloramphenicol that completely inhibited growth was designated as the MIC.

Random mutagenesis and selection. Random mutagenesis of the 790 loop was performed essentially by the method of Higuchi (1989) using PCR and cloned in pRNA122 using the unique *Bgl*II and *Dra*III restriction sites (Higuchi, R. (1989) PCR Technology (Erlich, H.A., ed.), pp. 61-70, Stockton Press, New York) (Figure 18). For each set of mutations, four primers were used: two "outside" primers and two "inside" primers. The two outside primers were designed to anneal to either side of the *Bgl*II and *Dra*III restriction sites in pRNA122 (*Figure 2*). These primers were 16S-*Dra*III, 5' - GACAATCTGTGTGAGCACTA-3' and 16S-535, 5' - TGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGT-3' . The inside primers were 16S-786R, 5' -CCTGTTTGCTCCCCACGCTTTCGCACCTGAGCG-3' and 16S-ASS-3, 5'- CTCAGGTGCGAAAGCGTGGGGAGCAAACAGGNNNNNNNNCCTGGTAGTCC ACGCC GTAA-3' (N = A, T, C and G). Thus, $4^9 = 262{,}144$ possible combinations were created, with the exception of 320 sequences that were eliminated because they formed either *Bgl*II or *Dra*III recognition sites (256 *Bgl*II sites and 64 *Dra*III sites).

Transformants were incubated in SOC medium containing 1 mM IPTG for four hours to induce rRNA synthesis and then plated on LB agar containing 100 μg/ml chloramphenicol. A total of 2×10^6 transformants were plated yielding approximately 2000 chloramphenicol-resistant survivors. Next, 736 of these survivors were randomly chosen and assayed to determine the MIC of chloramphenicol necessary to completely inhibit growth in cells expressing mutant ribosomes. From this pool, 182 transformants with MICs greater than 100 μg/ml were randomly selected and sequenced.

Site-directed mutation of positions 787 and 795. Mutations at positions 787 and 795 were constructed as described above for the random mutants, except that the inside primers were 16S-786R (see above) and 16S-ASS-4, 5' - CTCAGGTGCGAAAGCGTGGGGAGCAAACAGGNTTAGATANCCTGGTAGTCC ACGCCGTAA-3' (N = A, T, C and G). Transformants were selected on LB-Ap100

agar plates and grouped according to their MICs for chloramphenicol. Representatives from each group were then sequenced to identify the mutations.

Primer extension. To determine the ratio of plasmid to chromosome-derived rRNA, 30S and 70 S ribosomes were isolated from 200 ml of induced, plasmid containing cells by the method of Powers & Noller (1991). The purified RNA was then used in primer extension experiments (Triman, K., *et al.* (1989) *J. Mol. Biol.* 209:643-653). End-labeled primers complementary to sequences 3' to the 788 and 795 mutation sites were annealed to rRNA from induced cells and extended through the mutation site using AMV reverse transcriptase. The primers used were: 16S-806R, 5' -GGACTACCAGGGTATCT-3' ; 16S-814R, 5' -TACGGCGTGGACTACCA-3'. For wild-type pRNA122 ribosomes, position 1192 in the 16S RNA gene was changed from C to U and primers were constructed as described above (Triman *et al.*, 1989, *supra*). This mutation has previously been shown not to affect subunit association (Sigmund, C.D., *et al.* (1988) *Methods Enzymol.* 164:673-689). The extension mixture contained a mixture of three deoxyribonucleotides and one dideoxyribonucleotide. The cDNAs were resolved by PAGE and the ratios of mutant to non-mutant ribosomes were determined by comparing the amount of radioactivity in each of the two bands.

Oligoribonucleotide synthesis. Oligoribonucleotides were synthesized on solid support with the phosphoramidite method (Capaldi, D. & Reese, C. (1994) *Nucl. Acids Res.* 22:2209-2216) on a Cruachem PS 250 DNA/RNA synthesizer. Oligomers were removed from solid support and deprotected by treatment with ammonia and acid following the manufacturer's recommendations. The RNA was purified on a silica gel Si500F TLC plate (Baker) eluted for five hours with *n*-propanol/ammonia/water (55:35:10, by vol.). Bands were visualized with an ultraviolet lamp and the least mobile band was cut out and eluted three times with 1 ml of purified water. Oligomers were further purified with a Sep-pak C-18 cartridge (Waters) and desalted by continuous-flow dialysis (BRL). Purities were checked by analytical C-8 HPLC (Perceptive Biosystems) and were greater than 95%.

Experimental Procedures

Sequence analysis of functional mutants. Random mutations were introduced simultaneously at all nine positions (787 to 795) in the 790 loop. Functional (chloramphenicol-resistant) mutants were then selected in *E. coli* DH5 cells (Hanahan, 1983, *supra*) and the effects of these mutations on ribosome function were determined. A total of 182 mutants that retained chloramphenicol resistance were randomly selected and sequenced. Wild-type 790-loop sequences were obtained from 81 of the sequenced

transformants, while the remaining 101 contained mutant sequences. One of the transformants was chloramphenicol-resistant in the absence of inducer, presumably due to a spontaneous mutation in the CAT gene, and was excluded from further analysis. Of 100 sequenced functional mutants, 14 were duplicates and four sequences occurred three times. Thus, 78 different, functional, 790-loop mutants were analyzed (Figure 19). According to resampling theory, this distribution indicates that of the $4^9 = 262,144$ possible sequences, only 190 (standard deviation 30) unique sequences exist in the pool of selected functional mutants. Of the 78 mutants, 44 contained four to six substitutions out of the nine bases mutated and 21 of these retained greater than 50% of the wild-type activity. The minimal inhibitory concentration (MIC) of chloramphenicol for cells expressing wild-type rRNA from pRNA122 is 600 μg/ml. MICs of the mutants ranged from 150 to 550 μg/ml with a mean of 320 μg/ml (standard deviation 89). The median and mode were both 350 μg/ml.

Functional 790-loop mutants showed strong nucleotide preferences at all mutated positions, except positions 788 and 792, which showed a random distribution (Figure 20) but significant covariation. No mutations were observed at U789 or G791. Mutations at these positions, however, were present in mutants that were selected for loss of function (not shown). Thus, these nucleotides appear to be directly involved in ribosome function. U789 is strictly conserved among bacteria but is frequently C789 among other organisms (Figure 20). Chemical protection studies have shown that G791 is specifically protected from kethoxal modification in 70 S ribosomes and polysomes (Brow, D. A. & Noller, H. F. (1983) *J. Mol. Biol.* 163: 112-118; Moazed, D. & Noller, H.F. (1986) J. Mol. Biol. 191: 483-493); and by poly(U) (Moazed & Noller, 1986, *supra*) and that G791 becomes more accessible to kethoxal modification when 30S subunits are converted from the "inactive" to "active" conformation (Moazed *et al.*, 1986, *supra*).

Purines were strongly selected at position 787 (97.4%) while A and, to a lesser extent, C were preferred at position 790 (98.7%) and U was completely excluded at both positions. At both position 793 and 795, A, C and U were equally distributed but G was selected against. Adenine and uracil were preferred at position 794 (81.8%).

Non-random distribution of nucleotides among the selected functional clones indicates that nucleotide identity affects the level of ribosome function. To examine this, the mean activities (MICs) of ribosomes containing all mutations at a given position were compared by single-factor analysis of variance between ribosome function (MIC) and nucleotide identity at each mutated position. Positions that showed a significant effect of nucleotide identity upon the level of ribosome function were 787 ($P < 0.001$), 788 ($P < 0.05$) and 795 ($P < 0.001$). The absence of mutations at positions U789 and

G791 in the functional clones prevents statistical analysis of these positions but mutations at these positions presumably strongly affect ribosome function as well.

Figure 20 shows a comparison of the selected functional mutants with current phylogenetic data (R. Gutell, unpublished results; Gutell, R. R. (1994) *Nucl. Acids Res.* 22(17): 3502-3507; Maidak, B. L. *et al.* (1996) *Nucl. Acids Res.* 24: 82-85). While nucleotide preferences in the selected mutants are similar to those observed in the phylogenetic data, the mutant sequences selected in this study show much more variability than those found in nature. This may be because all of the positions in the loop were mutated simultaneously, allowing normally deleterious mutations in one position to be compensated for by mutations at other positions, a process that is unlikely to occur in nature. In addition, none of the mutants was as functional as the wild-type, suggesting that wild-type 790-loop sequences have been selected for optimal activity or that other portions of the translational machinery have been optimized to function with the wild-type sequence.

To identify potential nucleotide covariation within the loop, the paired distribution of selected nucleotides was examined for goodness of fit. The most significant covariations were observed between positions 787 and 795 ($P < 0.001$) and between positions 790 and 793 ($P < 0.001$). For positions 790 and 793, only eight double mutants were available for analysis; therefore, the covariation observed between these positions should be regarded with caution. Position 788, which showed no nucleotide specificity, did show significant covariation with positions 787 ($P < 0.01$), 794 ($P < 0.01$) and 795 ($P < 0.01$).

Analysis of site-directed mutations constructed at the base of the loop: Functional analysis of mutations at positions 787 and 795. The observed covariations among positions 787, 788 and 795 are particularly interesting, since nucleotide identity at these positions correlated with the level of ribosome function. Further analysis of nucleotides at positions 787 and 795 revealed that 72 of the 78 functional mutants have the potential to form mismatched base-pairs (A · C, G · U, A · A and G · G). Other mismatches, such as G · A and U · G, however, were not found. In addition, only four sequences with an A · U Watson-Crick pair and no sequences with a U · A, G · C or C · G pair were present, suggesting that strong base-pairs between these positions inhibit ribosome function. Therefore all possible nucleotide combinations at positions 787 and 795 were constructed and analyzed without changing other nucleotides in the 790 loop. Ribosome function of the mutants (Figure 21) varied from 84% (A · A) to 1% (C · G) of the wild-type. As predicted by analysis of the pool of functional

random mutants, site-directed mutants with G · C, C · G and U · A Watson-Crick pairs between positions 787 and 795 were strongly inhibitory.

Results

These data suggest that strong pairing between nucleotides at positions 787 and 795 inhibits ribosome function. In addition, some of the site-directed substitutions at positions 787 and 795 that produced functional ribosomes were largely excluded from the pool of mutants in which all of the loop positions were mutated simultaneously (*e.g.* CC, CU, UU and UC). The observed nucleotide preferences at positions 787 and 795 in the selected random pool presumably reflect interaction of nucleotides at these positions with other nucleotides in the loop. This is consistent with our findings of extensive covariations among these sites.

Perturbations of the 790 loop have been shown to affect ribosomal subunit association (Herr, W., *et al.* (1979) *J. Mol. Biol.* 130: 433-449; Tapprich, W. & Hill, W., (1986) *Proc. Natl Acad. Sci. USA* 83: 556-560; Tapprich, W., *et al.* (1989) *Proc. Natl Acad. Sci. USA* 86: 4927-4931). Therefore several of the 787 to 795 mutants were tested for their ability to form 70 S ribosomes. Ribosomes were isolated from selected mutants and the distribution of mutant ribosomes in both the 70 S and 30S peaks was determined by primer extension (Figure 21). These data show that CAT activity correlates with the presence of mutant 30S subunits in the 70 S ribosome pool. Thus, loss of function may be due to the inability of mutant 30S and 50 S subunits to associate. Another explanation for this observation is that the mutations may directly affect a stage of the protein synthesis process prior to subunit association, such as initiation, which prevents subsequent steps from occurring. Other mutations in the 16S rRNA have been identified for which this appears to be the case (Cunningham, P., *et al.* (1993) *Biochemistry* 32: 7172-7180).

The references cited in Example 5 may be found in Lee, K. *et al.*, *J. Mol. Biol.* 269: 732-743 (1997), expressly incorporated by reference herein.

EXAMPLE 6: CONSTRUCTION OF A HYBRID CONSTRUCT

A plasmid construct of the present invention identified as the hybrid construct, is set forth in Figures 17 and 25. This hybrid construct contains a 16S rRNA from *Mycobacterium tuberculosis.* The specific sites on the hybrid construct are as follows: the part of rRNA from *E. coli* rrnB operon corresponds to nucleic acids 1-931; the part of 16S rRNA from *Mycobacterium tuberculosis* rrn operon corresponds to nucleic acids 932-1542; the 16S MBS GGGAU corresponds to nucleic acids 1536-1540; the terminator T1 of *E. coli* rrnB operon corresponds to nucleic acids 1791-1834; the

terminator T2 of *E. coli* rrnB operon corresponds to nucleic acids 1965-1994; the replication origin corresponds to nucleic acids 3054-2438; the *bla* (β-lactamase; ampicillin resistance) corresponds to nucleic acids 3214-4074; the GFP corresponds to nucleic acids 5726-4992; the GFP RBS (ribosome binding sequence) AUCCC corresponds to nucleic acids 5738-5734; the trp^c promoter corresponds to nucleic acids 5795-5755; the trp^c promoter corresponds to nucleic acids 6270-6310; the CAT RBS (ribosome binding sequence) AUCCC corresponds to nucleic acids 6327-6331; the *cam* (chloramphenicol acetyltransferase; CAT) corresponds to nucleic acids 6339-6998; the $lacI^q$ promoter corresponds to nucleic acids 7307-7384; the $lacI^q$ (lac repressor) corresponds to nucleic acids 7385-8467; and the *lac UV5* promoter corresponds to nucleic acids 8510-8551.

All references cited herein are expressly incorporated by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.